



AFRL-RH-WP-TR-2015-0092

Physiological Indicators of Workload in a Remotely Piloted Aircraft Simulation

**Michael Hoepf
Oak Ridge Institute for Science and Education**

**Matt Middendorf
Middendorf Scientific Services**

**Samantha Epling
Ball Aerospace**

**Scott Galster
Air Force Research Laboratory**

October 2015

Interim Report

Distribution A: Approved for public release.

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-TR-2015-0092 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//

Kevin D. Schmidt
Work Unit Manager
Applied Neuroscience Branch

//signed//

Scott M. Galster
Chief, Applied Neuroscience Branch
Warfighter Interface Division

//signed//

William E. Russell
Warfighter Interface Division
Human Effectiveness Directorate
711 Human Performance Wing

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YY) 2015-10-07		2. REPORT TYPE Interim Report		3. DATES COVERED (From - To) 17 October 2014 – 1 October 2015	
4. TITLE AND SUBTITLE Physiological Indicators of Workload in a Remotely Piloted Aircraft Simulation				5a. CONTRACT NUMBER FA8650-14-D-6501-0009	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) **Michael Hoepf, +Matt Middendorf, #Samantha Epling, and *Scott Galster				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER H0HJ (53290813)	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **ORISE 4692 Millennium Drive, Suite 101Belcamp, Maryland 21017 ,+Middendorf Scientific Services, #Ball Aerospace & Technologies 2875 Presidential Drive Fairborn OH 454324				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory* 711Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Applied Neuroscience Branch Applied Adaptive Aiding Section Wright-Patterson Air Force Base, OH 45433				10. SPONSORING/MONITORING AGENCY ACRONYM(S) 711 HPW/RHCP	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RH-WP-TR-2015-0092	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A: Approved for public release.					
13. SUPPLEMENTARY NOTES 88 ABW Cleared 02/10/2015; 88ABW-2016-0538. Report contains color.					
14. ABSTRACT Toward preventing performance decrements associated with mental overload in remotely piloted aircraft (RPA) operations, the current research investigated the feasibility of using physiological measures to assess cognitive workload. Two RPA operators were interviewed to identify factors that impact workload in target tracking missions. Performance, subjective workload, cortical, cardiac and eye data were collected. One cardiac and several eye measures were sensitive to changes in workload as evidenced by performance and subjective workload data. This research advances the literature toward real-time workload mitigation in RPA field operations.					
15. SUBJECT TERMS Cognitive workload, cortical measures, cardiac measures, eye tracking, and physiological workload.					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 44	19a. NAME OF RESPONSIBLE PERSON (Monitor) Kevin D. Schmidt 19b. TELEPHONE NUMBER (Include Area Code)
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			

TABLE OF CONTENTS

Section	Page
List of Figures	iii
List of Tables	iii
ACKNOWLEDGMENTS	iv
1.0 SUMMARY	1
2.0 INTRODUCTION	1
2.1 Workload.....	3
2.2 Cortical Measures.....	3
2.3 Cardiac Measures	4
2.4 Eye Tracking	4
2.5 Remotely Piloted Aircraft (RPA).....	5
3.0 METHODS	5
3.1 Participants	5
3.2 Apparatus and Measures	6
3.3 Procedure.....	11
4.0 RESULTS	12
4.1 Performance	14
4.2 Subjective Workload.....	14
4.3 Cortical Measures.....	15
4.4 Cardiac Measures	16
4.5 Eye Measures	17
5.0 DISCUSSION	17
5.1 Limitations	19
5.2 Implications and Future Research	19
6.0 CONCLUSIONS.....	20
7.0 REFERENCES	21
APPENDIX A - Screenshots	28
APPENDIX B - EEG Reference.....	36
LIST OF ABBREVIATIONS AND ACRONYMS	38

LIST OF FIGURES

	Page
Figure 1 Timeline of key events. The number in parenthesis indicates the trial time in seconds, and the darker blocks are events associated with a second HVT, which were present in half of the trials.....	6
Figure 2 Interaction between number of HVTs and route on performance (+SE)	14
Figure 3 Saccade amplitude (+SE) for country and city conditions for each participant. Amplitude values are absolute values and in arbitrary units.....	16

LIST OF TABLES

	Page
Table 1 2x 2 x 2 Factorial Design.....	11
Table 2 Means, SEs (in parenthesis), and ANOVA results (F values and probabilities for the comparison between easy and difficult task levels) for dependent variables (excluding EEG)...	13

ACKNOWLEDGMENTS

The authors would like to thank Chelsey Credlebaugh and Jonathan Mead for their assistance in data reduction and preparation of this manuscript, as well as Chuck Goodyear for his help with data analysis. We would also like to thank Kevin Durkee, Noah DePriest, and Mark Squire for their technical support. This research was supported in part by an appointment to the Student Research Participation Program at the U.S. Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Interface Division, Applied Neuroscience Branch administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USAFRL.

1.0 SUMMARY

This report describes an experiment conducted in the Human Universal Measurement and Assessment Network (HUMAN) Laboratory. Toward preventing performance decrements associated with mental overload in remotely piloted aircraft (RPA) operations, the current research investigated the feasibility of using physiological measures to assess cognitive workload. Two RPA operators were interviewed to identify factors that impact workload in target tracking missions. Performance, subjective workload, cortical, cardiac and eye data were collected. One cardiac and several eye measures were sensitive to changes in workload as evidenced by performance and subjective workload data. The report also contains a discussion of physiological data processing techniques and challenges (i.e., eye-artifacts and the interpretation of cortical measures), and suggestions for future research. Potential applications of this research include closed loop systems that employ advanced augmentation strategies, such as adaptive automation. Thus, by identifying physiological measures well suited for monitoring workload in a realistic simulation, this research advances the literature toward real-time workload mitigation in RPA field operations.

2.0 INTRODUCTION

U.S. armed forces are increasing using RPA to accomplish missions in hostile territory. One proposal to accomplish more with less is to allow operators to control multiple aircraft simultaneously (Rose, Arnold, & Howse, 2013). However, piloting one aircraft remotely is a complex task, and operating additional aircraft could increase task demands sharply. This is potentially problematic because most people perform their best within an optimal range of cognitive workload. That is, both cognitive underload and overload can negatively impact performance (Young & Stanton, 2002).

One solution to offset this risk, though rarely implemented, is to monitor operator workload and provide augmentation before performance decrements occur (Wilson & Russell, 2007). Physiological measures, which have been shown to reflect changes in cognitive demand in various aviation environments (e.g., Christensen & Estep, 2013; Roscoe, 1992; Veltman & Gaillard, 1998; Wilson & Russell, 2007), are well suited for this goal. However, before physiologically based workload monitoring systems can be adopted into RPA field operations, it is necessary to conduct testing in the lab using ecologically valid task environments. This is especially true as future RPA control stations might look much different than current systems, for example, integrating pilot and sensor duties to allow the simultaneous control of multiple aircraft. The current research aims to advance the current state of the art by investigating the utility of physiological measures for monitoring workload in a futuristic RPA control station.

Researchers are continually exploring methods for assessing workload in real-time. Using functional magnetic resonance imaging (fMRI), researchers have shed significant light on the functioning of the human brain (Graham et al., 2010; Muller-Plath, 2008; Pouliot et al., 2012; Ye & Zhou, 2009). However, the current research utilizes non-invasive sensors, which can be worn while sitting and working on a computer. The understanding of workload has been advanced by researchers using a variety of non-invasive physiological measures such as electroencephalogram

(EEG; Berka & Levendowski, 2007; Gevins, DuRousseau, Zhang, & Libove, 1993; Gevins, Smith, McEvoy, & Yu, 1997; Grimes, Tan, Hudson, Shenoy, & Rao, 2008; Lotte, Congedo, L'ecuyer, Lamarche, & Arnaldi, 2007; Wolpaw, McFarland, Neat, & Forneris, 1991), electrocardiogram (ECG; Jorna, 1992; Mulder, 1992; Porges & Byrne, 1992; & Roscoe, 1992) and eye tracking (Iqbal, Zheng, & Bailey, 2004; Jacob, 1991; Marshall, Pleydell-Pearce & Dickson, 2003). These researchers have demonstrated the feasibility of real-time workload assessment; however, there are several areas in which the field needs to be advanced before these methods can be applied to RPA operations.

First, there is a need for research in realistic RPA task environments. Researchers often use conventional laboratory tasks to study physiological indicators of workload, such as Kirchner's (1958) *n*-back (e.g., Grimes et al., 2008), or Comstock and Arnegard's (1992) Multi-Attribute Task Battery (e.g., Christensen, Estepp, Wilson, & Russell, 2012). Such tasks afford researchers experimental control over specific components of workload. For example, by utilizing the *n*-back experimental paradigm, researchers can effectively manipulate working memory demand. While demonstrating that physiological measures have utility for monitoring workload in standard laboratory tasks is an important step, researchers have recommended that further sensor evaluation occur in realistic task environments (e.g., Zarjam, Epps, & Lovell, 2012). To implement physiologically based workload monitoring into field RPA operations, researchers will first need to show that such systems work in realistic RPA tasks in a laboratory setting.

Second, there is a need for research that examines which physiological measures are best suited for RPA tasks. Each physiological measure can provide unique information, redundant information, or no information about workload. Before such technology can be implemented into the field, bulky/redundant sensors will need to be eliminated. That is, some sensors are better suited for some situations than others. For example, Hankins and Wilson (1998) found that only eye activity was related to workload during visually demanding flight segments, whereas heart rate (HR) was related to workload during flight segments heavily dependent on instrument use, and EEG demonstrated sensitivity to mental calculation. This finding is consistent with Wickens' (1984) multiple resource theory (MRT), which distinguishes visual, auditory, tactile, and olfactory sources of input, processing, and action. We designed and implemented an experiment to address the research needs outlined above. The contributions of this research are two-fold:

To address the first need, we utilize a high-fidelity next generation multiple aircraft management RPA simulator to examine workload under various conditions. To make the task as ecologically valid as possible, we interviewed two RPA subject matter experts (SMEs) to identify high workload situations. We then implemented experimental manipulations to control those situations. We also examine the operation of one vs. two RPA.

To address the second need, we employ an array of physiological measures, including EEG, ECG, and eye tracking (including electrooculogram (EOG)). This allows us to simultaneously evaluate the effectiveness of multiple measures within the same task environment. Our goal is to identify physiological measures which are well suited for monitoring workload in RPA task environments.

The rest of this paper proceeds as follows. First, we review the relevant literature on workload, physiological measures, and RPA. Next, we describe an experiment designed to enhance the current body of literature as outlined above. We conclude with a discussion of our findings, limitations and suggestions for future research.

2.1 Workload

Cognitive workload is an easy concept to understand generally (how hard the brain is working), but because it is multidimensional and context dependant, it is difficult to define precisely (Hart & Staveland, 1988). According to resource based models of workload (e.g., Wickens, 1984), overload occurs when the demands of a task exceed an operator's ability to meet them. Considering both the operator's cognitive capacity and task demands is important, as workload can be viewed as an interaction between the two (Cain, 2007). An operator's cognitive capacity may change due to training, fatigue or the environment, while task difficulty may vary due to situational changes or task reallocation.

To reduce workload, it is necessary to accurately measure workload and identify the cause of the high workload. Researchers have recognized that subjective ratings, performance measures, and physiological measures can be valid indicators of workload (Hart & Staveland, 1988; O'Donnell & Eggemeier, 1986; Wilson et al., 2004). Subjective measures can be introduced at any time, but they may intrude into the operator's task (Kramer, 1991; Wilson & Eggemeier, 1991). We doubt that pilots would appreciate being prompted for a workload assessment during critical mission phases. Furthermore, if the collection of subjective measures is postponed to avoid this interference, operator responses might suffer from memory lapses and bias (Moroney, Biers, & Eggemeier, 1995).

Performance can be obtained continuously in some contexts, but, in jobs that are highly automated, operators are primarily in a monitoring role. This significantly reduces opportunities to observe performance (Wilson & Eggemeier, 1991). In addition, if the goal of monitoring workload is to prevent performance decrements, using performance to ascertain workload would clearly be insufficient in some situations. That is, performance decrements will have to occur *before* high workload can be recognized. It would not benefit commanders to recognize, for instance, that a pilot had experienced high workload after mission failure had already occurred.

Physiological measures can be continuous, nonintrusive, and have been shown to be sensitive to changes in mental and physical workload (e.g., Caldwell, Caldwell, Brown, & Smith, 2004; Wilson et al., 2004). Furthermore, physiological measures can be combined to provide a superior estimation of workload in complex, multifaceted tasks, such as air traffic control (Brookings, Wilson, & Swain, 1996) and piloting aircraft (Hankins & Wilson, 1998). Thus, physiological measures appear to be ideally suited for continuous workload assessment in RPA operations.

2.2 Cortical Measures

There are numerous neuroimaging techniques available for studying the complex and dynamic behavior of the brain. However, due to prohibitive factors such as cost and portability, the current study uses electroencephalography (EEG). EEG is the recording of electrical activity

along the scalp, which measures voltage fluctuations resulting from ionic current flows within the neurons of the brain (Niedermeyer & da Silva, 2004). Advantages of EEG include high temporal resolution, ease of use, and relatively low cost compared to other neuroimaging techniques (Zander & Kothe, 2011). Typical methods to examine EEG data include: power spectral density or the averaged power, maximum/log power spectra, sub-band entropy, and autoregressive modeling (Zarjam et al., 2012). Researchers have demonstrated that EEG can be used in real-time to assess mental workload (e.g., Wilson & Russell, 2007), and that such methods are sufficiently stable to provide accurate assessment over the course of several days and weeks (Christensen et al., 2012).

2.3 Cardiac Measures

Researchers have investigated the relationship between cardiac measures and workload for many years. Such measures are relatively easy to obtain and can be assessed continuously in most task environments. Typical cardiac measures include heart rate (HR) and heart rate variability (HRV). These measures can be acquired with the application of a few electrodes over the heart. If HR is all that is required, it may suffice to simply place a monitoring device where artery pulsation is transmitted to the surface, such as the wrist. Measures of HR have the longest history and the widest use, while fewer researchers have used HRV (Wilson, 1992). In both laboratory and field settings, researchers typically observe HR increases and HRV decreases in high workload situations (e.g., Jorna, 1992; Mulder, 1992; Porges & Byrne, 1992; & Roscoe, 1992). However, there is some debate about which measure is superior. Roscoe (1992) suggested that HRV may indicate changes in mental workload in the absence of any change in overall HR. Yet, Grossman (1992) indicated that it is not clear that HRV provides any more information than simple HR.

Despite the ease of collecting cardiac measures, there are potential obstacles associated with their use. Wilson (1992) observed that participants are often unfamiliar with the task and unaccustomed to being in a laboratory setting. Thus, learning effects may account for some of the variance in cardiac measures in laboratory tasks. However, applied settings present a lack of experimental control. Differences in HR are often substantially larger in applied settings than in laboratory settings (Wilson, 1992). Changes in cardiac activity might be smaller in laboratory setting than field settings because there are no dire consequences as a result of poor performance in laboratory tasks. Also, researchers need to be cautious not to confuse changes in cardiac measures that arise from cognitive activity with changes due to physical activity (Wilson, 1992). Finally, Roscoe (1992) noted that HR can be affected by environmental factors such as heat, vibration and noise, by stimulants (e.g., caffeine and tobacco), by pain and discomfort, and by diurnal variations. It is worth noting that, there are limitations with all physiological measures, and we would expect many of the obstacles discussed here to apply to other measures as well.

2.4 Eye Tracking

Eye tracking is a general term covering a variety of eye-based measures. Physiological workload features that can be derived from eye tracking are typically categorized into four groups: blinks, saccades, fixation, and pupillary response. We focused on blinks and pupillary measures in the current research. Regarding blinks, Wang and Zhou (2013) concluded that: 1. Blink rate

decreases with an increase in cognitive load, and 2. Blink duration tends to decrease during more intense processing load.

Pupil diameter fluctuates based on both lighting and autonomic nervous system responses (e.g., the fight or flight response of the sympathetic nervous system in a threatening situation causes pupil dilation). An increase in pupil diameter can be associated with an increase in mental demand (Beatty, 1982; Wang & Zhou, 2013). However, pupil dilation changes from the illumination condition of the visual field can be strong. Pomplun and Sunkara (2003) observed that background brightness resulted in greater variation of pupil diameter than task difficulty.

In sum, cortical, cardiac, and eye based physiological measures are all potentially well suited for monitoring workload. Thus, data will be collected from each of these sources in the current research. We now turn the discussion to RPA operations.

2.5 Remotely Piloted Aircraft (RPA)

RPA are in high military demand because of their standoff capability in areas that are difficult to access or otherwise considered too hazardous for manned aircraft or personnel on the ground (U.S. Department of Defense, 2011). It has been documented that the military intends to increase the number of RPA in service while simultaneously reducing the number of operators (Dixon et al., 2004). It is envisioned that the next generation RPA will have single human operators monitoring multiple semi-autonomous RPA. The benefits of such automated systems include: reducing manpower requirements, lower life-cycle costs, and decreased human exposure to hazardous environments (Prabhala, Gallimore, & Narayanan, 2003; Ruff, Narayanan, & Draper, 2002). Such systems would require operators to perform high-level cognitive tasks such as: coordinating multiple RPA, overseeing multiple target areas, detecting targets, identifying targets, route planning, and monitoring system status (Liu, Wasson, & Vincenzi, 2009). Thus, there is a need to minimize the mental workload demands that will be imposed on the operators of such systems. A reduction in workload should enable better overall system performance (Tsang & Wilson, 1997) and contribute to the reduction of the operator-to-vehicle ratio. In addition, we expect a reduction in mistakes such as equipment loss and ground casualties may eventually be possible. The following experiment investigates the utility of several physiological measures for monitoring workload in a futuristic RPA system.

3.0 METHODS

3.1 Participants

Participants in this study consisted of six people who were either students at a Midwestern university, or recent graduates. They were paid \$15 per hour for their participation. The participants were screened for motor, perceptual, cognitive, heart, and neurological conditions, as well as hearing impairments. They did not take any neurological medications or medications that caused drowsiness. Participants were comfortable operating a computer, reading small characters on a computer monitor, hearing and comprehending verbal commands presented through headphones, and learning complex, computer based tasks. Three participants were female and

three were male. Age ranged from 19-28, with a mean of 22. They were fluent in English, right-handed, had normal or corrected-to-normal eyesight with no color blindness, and provided written informed consent in accordance with human research ethics guidelines prior to the start of the experiment. All study procedures were reviewed and approved by the Air Force Research Laboratory Institutional Review Board.

3.2 Apparatus and Measures

Primary Task. The experiment consisted of a primary and secondary task. The primary task was developed using an RPA software platform known as the Vigilant Spirit Control Station (VSCS; Rowe, Liggett, Davis, 2009). We utilized version 3.14, which was not updated or altered at any point during data collection. An important aspect of VSCS is that it allows one operator to control multiple RPA simultaneously. Our configuration allowed operators to control two RPA. Each RPA had one sensor, and we simulated the video feeds from these sensors using Virtual Reality Scene Generator (VRSG) software. The task was configured for a triple monitor control station such that the tactical situation display, which shows the map and aircraft information, was displayed on the left monitor, and the middle screen displayed the video feeds from the sensors. The right monitor displayed the communication windows. The primary task environment was centered on a simulated replica of a small town in Afghanistan, although the architecture of the town was randomly generated.

The goal of the primary task was to track one or two high value targets (HVTs). A timeline of the key events can be found in Figure 1. The first 30 seconds were designated as time for the participants to set up the RPA. This consisted of taking control of both RPA, adjusting the sensors, turning on the heads up display (HUD), and enabling sensor slaved tracking (an automation feature in which the aircraft flies auto-piloted loiter circles around the center of the sensor feed).

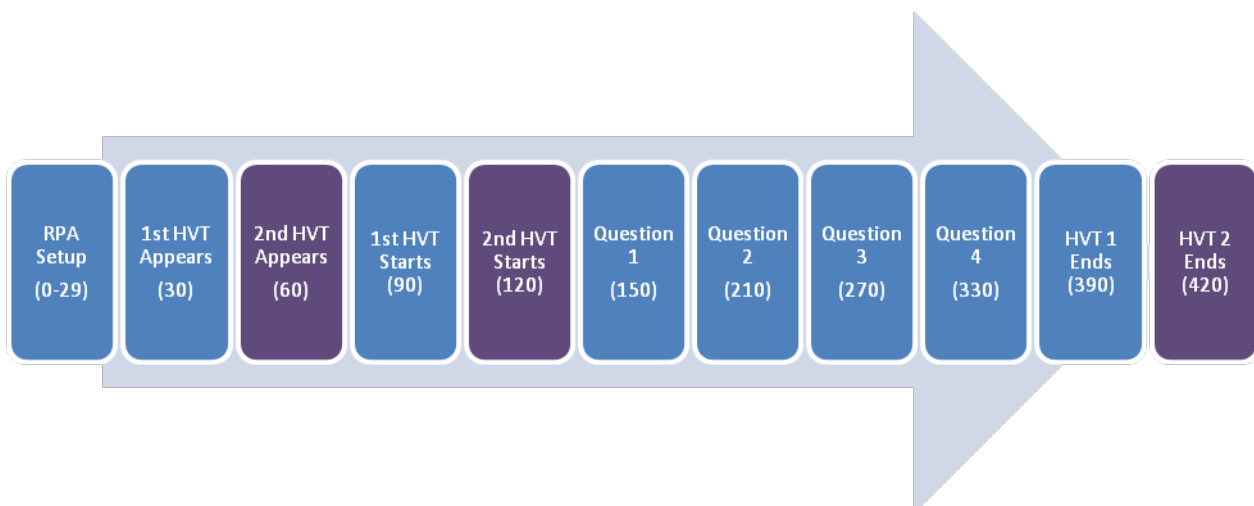


Figure 1. Timeline of key events. The number in parenthesis indicates the trial time in seconds, and the darker blocks are events associated with a second HVT, which were present in half of the trials

At 30 seconds into the trial, the first HVT walks out from underneath a tent and begins walking to a different tent. In half of the trials, at 60 seconds, a second HVT was initiated from the same tent as the first HVT. At 90 seconds, the first HVT would arrive at a second tent, and then leave on a motorcycle. If the trial had a second HVT, he would follow the same pattern, leaving the market via motorcycle 120 seconds into the trial. Subsequently, the HVT(s) would ride to a predetermined tent, either taking a route through the city or the country, depending on the condition. The routes took 300 seconds to complete, and thus trials with one HVT ran for 390 seconds, while trials with two HVTs lasted for 420 seconds.

Participants were instructed to keep the RPA sensor positioned over the HVTs, which they accomplished by clicking in the video feed with the mouse, causing the video feed to center on where they had clicked. The sensor slaved tracking feature would then automatically update the aircraft position to fly a loiter circle around this center point. This eliminated the need for the operator to manually navigate the aircraft. Participants were free to develop their own strategy regarding the location and frequency of where they clicked and to adjust the level of zoom of the sensor feed.

Secondary Task. The secondary task consisted of answering cognitively challenging questions. Questions were presented verbally over a headset, and were then made available visually. Participants were required to respond verbally by speaking into a headset. The secondary task was presented concurrently with the primary task.

There were three math questions and one mental rotation question per trial. Questions of the same type were presented at the same time in each trial, but were selected quasi-randomly (balanced for difficulty) from a bank of questions so participants could not anticipate specific questions. In each trial, a division question was presented at 150 seconds (e.g., "How long would it take you to reach a location 240 nautical miles away based on your current speed?"). A compound arithmetic question was presented at 210 seconds (e.g., "How long would it take you to reach a destination 100 nautical miles away with a headwind of 15 knots?). An addition or subtraction question was presented at 270 seconds (e.g., "What would your altitude be if you moved 1200 feet higher?") Finally, a mental rotation question was presented at 330 seconds, which was always the same ("What is the current heading of the HVT?"). For this question, participants had to look at the compass on the HUD and make a determination of which direction the first HVT was heading. This task was challenging because the viewing angle and compass were constantly rotating as the RPA flew loiter circles around the HVT. Furthermore, the heading question was unique in that it required participants to examine the video feed, whereas the first three questions could be calculated mentally because RPA speed and altitude were held constant and therefore easily memorized by the participant.

The secondary task was created using the Multi-Modal Communication (MMC) tool (Finomore, Popik, Dallman, Stewart, Satterfield, Castle, 2011). This software allowed scripted voice communications and transcriptions to play along a predetermined timeline. The MMC tool was set up on the right monitor, and was configured to contain two windows, one entitled "Mission", and one entitled "Response". Transcriptions of voiced questions were displayed in the mission window, and participants were required to respond by clicking and holding a push-to-talk button in the response window while they responded verbally.

Physiological Data Acquisition. The physiological data collected in this study included the EEG, ECG, vertical EOG (VEOG), and pupil diameter. The EEG data were acquired using electrodes placed directly on the scalp and secured in place with an Electro-Cap manufactured by Electro-Cap International, Inc. EEG was measured at seven sites on the scalp in accordance with the International 10/20 system (Jasper, 1958). The seven sites were the F7, F8, T3, T4, Fz, Pz, and O2 (see Appendix B). The right and left mastoids were used as the reference and ground for the EEG signals. All initial electrode impedances were measured to be at or below 5 k Ω .

The VEOG data were acquired using two electrodes placed above and below the left eye. The ECG data were acquired using two electrodes placed on the sternum and xiphoid process. The initial electrode impedances for the VEOG and ECG were measured to be at or below 20 k Ω . The left mastoid was used as the ground for the VEOG and ECG signal. The EEG and VEOG signals were sampled using two Cleveland Medical Devices BioRadio 150s at a sampling rate of 480 Hz. The ECG signal was sampled at 960 Hz. All signals connected to the BioRadio 150s were subjected to a hardware high pass filter with a break frequency of 0.5 Hz. The sampled data were transmitted wirelessly to a computer for processing and recording. The pupil diameter data were collected using the Smart Eye Pro 5.9 system, which included four cameras sampling data at 60Hz.

Physiological Data Processing. The raw VEOG signal was processed in real-time by a blink detection algorithm. The main features computed by the algorithm are blink duration and blink rate. The major components of the blink detection algorithm are threshold generation, a feature extraction state machine, and blink classification. The threshold generation component produced a threshold based on a sliding five second window of raw VEOG data. The feature extraction state machine finds the slope up, slope down, amplitude, and duration each time the VEOG signal goes above and below the threshold. The blink classification component compares these extracted features to criteria values to determine if the signal excursion above and below threshold is a blink.

The frequency bands (i.e., pass bands) used in the EEG signal processing are consistent with the traditional EEG bands. These bands are delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), gamma 1 (31-40 Hz), gamma 2 (41-57 Hz) and gamma 3 (63-100 Hz). A two second time domain window was used to process the raw EEG data. The raw data in the two second window was filtered using a 4th order Butterworth band pass filter. A Hanning window was applied to the filtered data and power spectral analysis was performed. The resulting power in the pass band was then averaged. These steps were repeated for each frequency band and electrode site. The two second time domain windows had a 50% overlap, thus yielding one measure of average power every second. This signal processing approach yielded 49 EEG measures per second (7 sites with 7 bands per site).

The two second window of raw EEG data was also processed to detect saccades. A unique feature of this EEG-based saccade detection is that the saccades are detected on a per-site basis. A linear fit with an initial length of 25 milliseconds is performed at the leading edge of the window. If the linear fit passes a slope and r^2 criteria, then the length of the fit is allowed to grow until the fit fails the criteria. The length, amplitude, and slope of the final fit is saved. The fit

length is then reset to 25 milliseconds and slid by one point towards the trailing edge of the window. The detection process is then repeated until the trailing edge of the window is reached. If one or more saccades are found in the window, the one with the largest amplitude is recorded and a flag is set to indicate that an artifact was found in the EEG window.

To account for EOG artifacts, the EEG data is examined in three ways. First, the results are reported with all of the EEG windows included. That is, no data is removed because of eye movement. Second, the EEG results are examined without windows that contain blinks. Third, results are examined without windows containing blinks or saccades.

The ECG data were processed by a peak detection algorithm. The inverse of the period between consecutive peaks was used to compute instantaneous HR frequency in hertz. The difference between consecutive HR frequencies was used to compute HRV in real-time.

The pupil diameter data from the Smart Eye Pro system was processed using logic to determine if the data was reliable. The system produces a reliability value for each pupil diameter measure, ranging from zero to one. The pupil diameter measures with a corresponding reliability value greater than 0.65 were used to compute an average pupil diameter measure for each trial. Lastly, all physiological data were inspected for outliers, which can be caused by sensor failure, participant movement, data transmission dropouts, and algorithm errors. Additionally, each of the EEG measures had an associated flag that would indicate the presence of an artifact (blinks and saccades). The blink indication came directly from the VEOG blink detection algorithm. The saccade indication came from a separate algorithm that was developed to directly detect saccades in the raw EEG data.

Composite Scoring Algorithm. Performance was measured using a composite scoring algorithm, which was based on components from both the primary and secondary task. For each trial, the maximum possible score was 1,000 points, (800 primary and 200 secondary). To obtain points on the primary task, participants were required to keep the HVT(s) in their video feed(s). Points would accrue at two different rates depending on the level of zoom. The maximum number of points (about 3 per second) was allotted when using the highest two levels of zoom (6.9x and 23.9x), whereas using the other levels (1.0x-4.4x) resulted in accruing half as many points per second (see Appendix A – Condition 4 for a screenshot of 6.9x and 4.4x). The rationale for this differential point structure was to better control workload. That is, using higher zoom levels resulted in higher levels of workload. Thus, the point structure rewarded participants for using higher levels of zoom (thereby experiencing greater workload) when it was possible for them to do so. In addition, we did not want participants to be able to simply zoom out so far that the HVTs were always in their video feed, thereby obtaining full points without any effort (i.e., cheating).

In the secondary task, participants had to answer questions correctly, and within a certain timeframe to acquire points. There were four questions per trial, each worth a maximum of 50 points. To obtain all 50 points, participants had to respond correctly within 10 seconds. After 10 seconds, the participants would lose 1 point per second for the next 10 seconds, and then 2 points per second for the next 10 seconds. After 30 seconds, no points were given. Answering incorrectly resulted in a 5 point penalty. However, an exception was made for the mental rotation question. We could not require an exact response because the HVTs frequently changed

direction, which meant that the HVT heading varied depending on when the participant answered the question. Thus, responses were considered correct for this question as long as a response was provided.

Competition. Individuals have a natural tendency to search for information in order to make comparisons of their behavior and attitudes with others (Festinger, 1954). This information often serves as a cue for competition (Seta, 1982). Competition has been defined as a use of social comparison information to make evaluative performance rankings (Martens, 1975). Furthermore, individuals have a tendency to challenge themselves in an effort to develop competence and prove self-determination (Deci & Ryan, 1991). So, when presented with performance feedback, people are motivated to continuously improve their own performance, and beat the performance of others. Thus, we implemented a “social” or “informal” competition to increase motivation and task engagement. The main difference between formal and informal is the focus of participant attention (Sommer, 1995). Informal contests direct attention towards behaviors and performance, and their comparisons. Conversely, formal competition directs the focus onto outcome issues such as prizes, allocation of scarce resources, and the justice of the allocation procedure. We offered no prizes for high scores, as the goal was for the competition to remain social. This is because social competitions have been associated with enhanced performance, whereas formal competition can lead to negative outcomes, such as higher intergroup hostility (Sherif, Harvey, White, Wood, & Sherif, 1961).

Another important aspect of competition is that it is associated with increased physiological reactivity. Cooke, Kavussanu, McIntyre, and Ring (2011), for instance, found that competition increased effort and heart rate while decreasing heart rate variability. Similarly, other researchers have observed an increase in heart rate and blood pressure with competition (Harrison et al., 2001; Veldhuijzen van Zanten et al., 2002). This could be seen as an advantage in the current research because we are looking for physiological changes associated with high workload. That is, we believe the inclusion of the competition will increase effort and physiological reactivity.

To implement the competition, top scores were posted on a whiteboard. This whiteboard was placed to the right of the participant control station so that it was out of view during the task, but easily visible during setup and between trials. The posted scores were averages based on all trials from each day. This encouraged the participants to remain engaged for all of the trials in the session. Participant identification numbers were used to maintain anonymity. The participants were debriefed following their final day of data collection. All twelve participants in the study indicated that the performance feedback positively affected their motivation, and the competition prevented task disengagement.

Subjective Workload. Subjective workload was assessed using the National Aeronautics and Space Administration Task Load Index (TLX), a multidimensional measure that assesses perceived workload (Hart & Staveland, 1988). Workload was determined by averaging across the six sub-scales (mental demand, physical demand, temporal demand, performance, effort, and frustration). Each scale was rated from 1-100 with the left anchors indicating “low” and the right anchors indicating “high” for all scales except the performance scale, which is reversed. Even though the TLX authors suggested a weighting procedure, we opted to use the simple average based on several findings in the existing literature. That is, the un-weighted average of the six

sub-scales has been found to be psychometrically (Nygren, 1991), and empirically equivalent to the weighted sub-scale averaging (Christ et al., 1993; Hendy, Hamilton, & Landry, 1993). In addition, there is some concern that the six sub-scales are often perceived as measuring only one or two constructs, and interpretation of the individual sub-scales should only be made with caution (Bailey & Thompson, 2001). Thus, the un-weighted average was used in the current research.

3.3 Procedure

Participants were brought into the laboratory for one day of training and four days of data collection. For training, participants first viewed a PowerPoint presentation containing a description of the primary and secondary tasks accompanied by screenshots. The presentation also contained information pertaining to subjective workload assessment and instructions for completing the TLX. Further, participants were read a script that encouraged them to accurately complete the TLX. The researchers then provided part-task training on the primary task and administered eight practice trials. Following each trial, participants completed a paper and pencil version of the TLX and were provided performance feedback.

On data collection days, researchers equipped the participants with the physiological devices, which took approximately 40 minutes to complete. The researchers then verified that the participants were comfortable, adjusted the volume of the headset, and provided several reminders regarding important components of the task. Participants completed eight experimental trials, separated by a break after the fourth trial. Baseline data was collected at the beginning and end of each session. Both baseline conditions consisted of five minutes of the participants passively viewing RPA video feeds. The four data collection days were structured the same except that a debriefing was conducted at the end of the fourth day.

Experimental Manipulations. The current investigation utilized a 2 x 2 x 2 full factorial design. There were three manipulations (summarized in Table 1) intended to impact workload, each containing two levels. The first manipulation was weather (i.e., visibility), which included clear and hazy conditions. The weather effects were created using the virtual reality scene generator. The clear condition was free of clouds and visibility was unobstructed. A layer of fog was present in the hazy condition, which reduced visibility. Workload was expected to be higher under hazy condition.

Table 1. 2 x 2 x 2 Factorial Design

	Clear		Hazy	
	Country	City	Country	City
One HVT	Condition 1	Condition 2	Condition 5	Condition 6
Two HVTs	Condition 3	Condition 4	Condition 7	Condition 8

The second manipulation was the number of HVTs, which was either one or two. In single HVT conditions, participants were required to track only one HVT, which meant it was only necessary

to use one RPA. In the two HVT condition, participants had to use both RPA to track the two targets. Workload was expected to be higher in two HVT condition.

The third manipulation was route difficulty, which refers to the roads the HVTs would travel on their motorcycles. The two route types were country and city. For the country routes, HVTs simply traveled back and forth along a long straight road, the view of which was generally not obstructed by any buildings. Conversely, for the city routes, the HVTs took many turns and sometimes became occluded by buildings. It was expected that city routes were more difficult than the country routes because of the turns and occlusion. Screenshots of each condition, as well as the task environment in general, are available in Appendix A.

Condition Balancing. When counterbalancing the order of the conditions, we were constrained by a lengthy virtual reality scene generator restart process required to load new weather settings. It was necessary to balance the order of the conditions within blocks of four trials, holding weather constant within each block. By doing so, we were only required to restart the software once during data collection, which also allowed for a consistent break time for participants half way through the session. Furthermore, each block of four trials was counterbalanced using balanced 4 x 4 Latin squares, which ensured that each treatment condition preceded every other treatment an equal number of times, thereby accounting for order effects (e.g., learning, fatigue, contrast).

To ensure that participants did not become overly familiar with the HVT routes and questions, we structured the experiment such that participants only experienced each HVT route and each question once per session. For the primary task, twelve HVT routes were created (four trials with one HVT and four trials with two HVTs). The number of turns and the amount of building obstruction was kept as consistent as possible within route type. Six routes were country, and six routes were city. Furthermore, three city routes were localized in one region of the town, whereas the other three stayed in a separate region. The country routes were similarly divided along a country road. Thus, we specifically structured the task such that multiple HVTs would never be in the same video feed at the same time. This was necessary to balance the difficulty of the conditions and avoid any unanticipated changes in workload associated with HVT confusion. In addition, we ensured that routes were counterbalanced such that each route appeared in clear and hazy conditions an equal number of times. To the extent possible, we also balanced the number of times that routes were used in single HVT conditions and dual HVT conditions. For the secondary task, there were four types of questions. We created a bank of eight questions for each question type, with the exception for the heading question. The heading question was always asked the same way, but this was acceptable because the expected response was different depending on the position of the RPA. Thus, as was the case with the routes, participants were exposed to each question once per session.

4.0 RESULTS

The performance, subjective workload, and physiological data were statistically evaluated using a three-way (weather, HVT, route) repeated-measures ANOVA. The means, standard errors (SE), and ANOVA results (excluding EEG) are presented in Table 2. EEG results are presented

separately (see Appendix B) because it is easier to view the large number of EEG measures via a visual depiction of the results over the scalp.

Table 2. Means, SEs (in parenthesis), and ANOVA results (F values and probabilities for the comparison between easy and difficult task levels) for dependent variables (excluding EEG)

	Weather	HVTs	Route
Performance			
Easy	776.2 (23.4)	873.6 (24.1)	814.5 (19.2)
Difficult	785.0 (25.6)	687.6 (25.4)	746.7 (31.6)
ANOVA	<i>ns</i>	$F(1,5) = 220.30, p < .001$	$F(1, 5) = 10.18, p < .05$
Subjective Workload (TLX)			
Easy	43.3 (5.0)	32.1 (4.2)	39.1 (4.1)
Difficult	43.5 (4.3)	54.6 (6.1)	47.6 (5.3)
ANOVA	<i>ns</i>	$F(1, 5) = 18.97, p < .01$	$F(1, 5) = 18.52, p < .01$
HR (beats per minute)			
Easy	70.51 (2.40)	69.99 (2.44)	70.76 (2.67)
Difficult	70.91 (2.55)	71.43 (2.61)	70.66 (2.29)
ANOVA	<i>ns</i>	<i>ns</i>	<i>ns</i>
HRV			
Easy	0.0533 (0.0046)	0.0555 (0.0050)	0.0542 (0.0047)
Difficult	0.0539 (0.0050)	0.0517 (0.0046)	0.0530 (0.0049)
ANOVA	<i>ns</i>	$F(1,5) = 19.46, p < .01$	$F(1,5) = 7.44, p < .05$
Blink Duration (s)			
Easy	0.1053 (0.0041)	0.1099 (0.0041)	0.1064 (0.0043)
Difficult	0.1051 (0.0044)	0.1005 (0.0047)	0.1041 (0.0042)
ANOVA	<i>ns</i>	$F(1,5) = 13.81, p < .05$	$F(1,5) = 16.77, p < .01$
Blink Rate (blink per minute)			
Easy	19.16 (5.08)	21.65 (5.87)	19.59 (5.23)
Difficult	18.77 (5.05)	16.28 (4.50)	18.34 (4.88)
ANOVA	<i>ns</i>	$*F(1,5) = 3.98, p = .10$	$F(1,5) = 8.23, p < .05$
Pupil Diameter (mm)			
Easy	4.06 (0.68)	3.84 (0.65)	3.92 (0.66)
Difficult	3.87 (0.59)	4.09 (0.62)	4.00 (0.63)
ANOVA	$\bullet F(1,5) = 14.85, p < .05$	$F(1,5) = 44.33, p < .01$	<i>ns</i>

Note. *ns* = not significant; * = near significant ($p = .10$); \bullet = means not in direction expected.

4.1 Performance

Performance in hazy conditions (784.97) was not significantly different than the performance in clear conditions (776.20). Performance was higher in conditions involving one HVT (873.56) compared to conditions involving two HVTs (687.60), $F(1,5) = 220.30$, $p < .001$, and higher in conditions with country routes (814.51) than in conditions with city routes (746.66), $F(1, 5) = 10.18$, $p < .05$.

In addition to the main effects on performance, there was an interaction between the number of HVTs and route type, $F(1, 5) = 13.73$, $p < .05$. The performance for conditions containing country routes declined from 882.88 in those with one HVT, to 746.31 in those with two, a difference of 136.57. By comparison, performance of conditions featuring city routes declined from 864.25 in conditions with one HVT, to 629.06 in those with two, a difference of 235.19 (see figure 2). This finding is not surprising given that when a second HVT was present in a trial, it always matched the route type of the first HVT. Adding a city HVT simply resulted in lower performance than adding a country HVT.

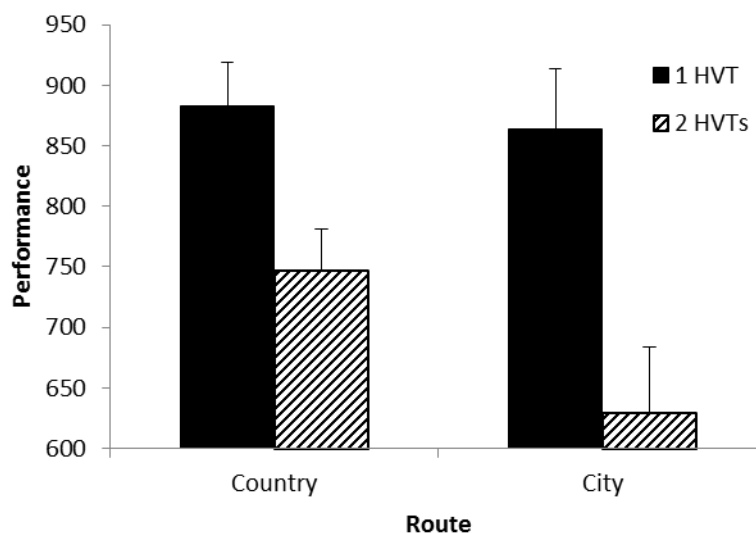


Figure 2. Interaction between number of HVTs and route on performance (+SE)

4.2 Subjective Workload

The workload of hazy conditions (43.48) was not significantly different than the workload of clear conditions (43.29). The workload of conditions involving two HVTs (54.64) was higher than the workload of conditions involving one HVT (32.13), $F(1, 5) = 18.97$, $p < .01$. The workload of conditions featuring city routes (47.63) was higher than conditions featuring country routes (39.14), $F(1, 5) = 18.52$, $p < .01$.

4.3 Cortical Measures

The EEG measures (power at each site and frequency band) were examined for each of the experimental manipulations. For conciseness in the text, we only report the results of the ANOVAs using all of the EEG data (no windows removed due to saccades or blinks), and using the EEG data with EOG artifacts removed (windows removed that contains blinks or saccades). A visual depiction of these results, as well as the EEG results based on the removal of blinks only, is available in Appendix B.

When all data was analyzed for the weather manipulation (clear vs. hazy), there was less power in the alpha band for the O2 site in the hazy condition than the clear condition $F(1, 5) = 9.24, p < 0.05$. When the artifact free data was analyzed for the weather manipulation, there was less power in the alpha band for O2 $F(1, 5) = 16.87, p < 0.05$ and PZ $F(1, 5) = 16.49, p < 0.05$ in the hazy condition than the clear condition. It is interesting that significance was found at an additional site when looking at the artifact free data. A decrease in alpha power is often associated with high workload (e.g., Dussault, Jouanin, & Guezennec, 2004; Prinz et al. 2003; Wilson, 2002).

When all data was analyzed for the route (county vs. city) manipulation, there was less power in the hard (city) condition than in the easy (country) condition in the delta band at F7 $F(1, 5) = 14.41, p < 0.05$, F8 $F(1, 5) = 8.49, p < 0.05$, T3 $F(1, 5) = 18.96, p < 0.05$, T4 $F(1, 5) = 9.83, p < 0.05$, Fz $F(1, 5) = 17.33, p < 0.05$, and Pz $F(1, 5) = 7.58, p < 0.05$, and in the theta band at F7 $F(1, 5) = 6.81, p < 0.05$. When the artifact free data was analyzed for the route (county vs. city) manipulation, there was more power in the theta band for the hard (city) condition than the easy (country) condition, the theta band at Fz $F(1, 5) = 27.51, p < 0.05$ and Pz $F(1, 5) = 25.09, p < 0.05$, and the gamma 1 band at T3 $F(1, 5) = 12.24, p < 0.05$. We believe that the significance in the delta and theta bands based on all of the EEG data is due to eye artifacts. EOG artifacts are a major source of concern in the interpretation of EEG data (Fatourechi, Bashashati, Ward & Birch, 2007). Furthermore, EOG activity has a wide frequency range, being maximal at frequencies below 4 Hz, and is most prominent over the anterior head regions (McFarland, McCane, David, & Wolpaw, 1997). This means that blinks and saccades can lead to an increase in power in the lower frequency EEG bands, especially delta. In the current investigation, participants blinked less during the hard (city) conditions (see EOG results), and saccade amplitude was greater in country conditions (see Figure 3). Thus, it appears that eye artifacts may explain why there was significantly less power at many sites in the delta band and one site in the theta band in hard (city) conditions. However, it is interesting to note that when the artifact free data was analyzed, more power was found at two sites in the theta band during hard (city) conditions. Theta power is often observed with an increase in workload (e.g., Hankins & Wilson, 1998; Wilson, Caldwell, & Russell, 2007), and thus this finding is not surprising.

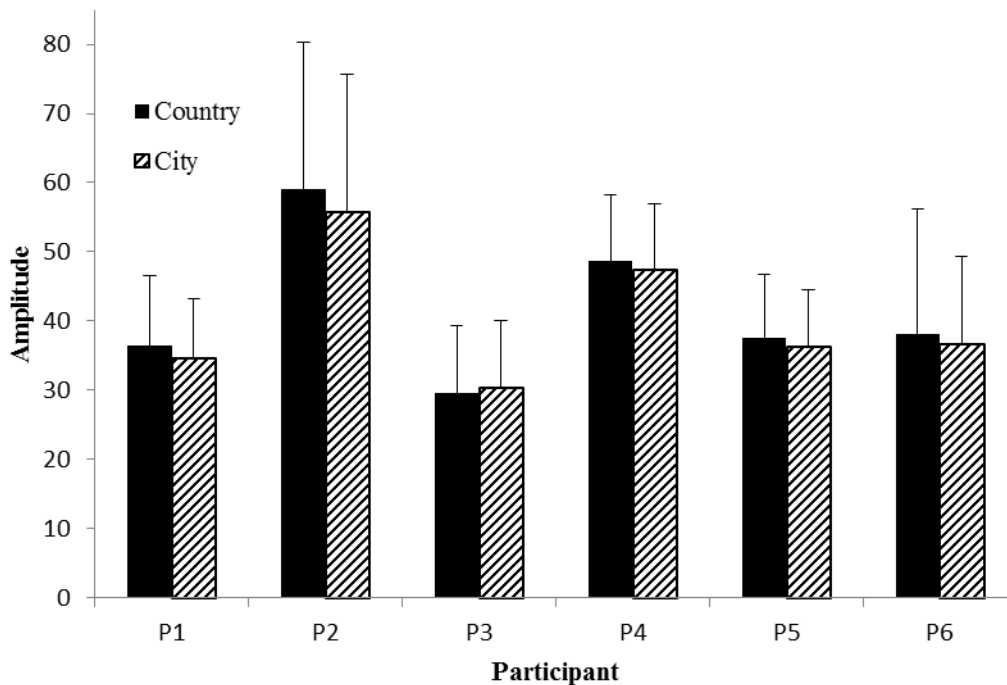


Figure 3. Saccade amplitude (+SE) for country and city conditions for each participant. Amplitude values are absolute values and in arbitrary units

When all EEG data was analyzed for the number of HVTs (one vs. two) manipulation, there was significantly more power ($p < .05$) for hard (2 HVT) conditions at several sites for the delta, theta, and alpha bands, and the T4 site for the beta, gamma 1, gamma 2, and gamma 3 bands. Interestingly, analyzing the artifact free data produced a similar pattern of results. In fact, an increase in power was found for an additional 7 measures across various sites and bands in the artifact free data vs. all of the data. This ‘across the board’ increase in power warranted further investigation. It appears that there is a task-related effect of eye movement that may explain this finding. When tracking two HVTs, the participant must constantly shift their gaze between two sensor feeds, thus introducing substantially more saccades. The large number of saccades would explain the increase in power when all of the EEG data were analyzed. However, one would expect that when the artifact free data was analyzed this effect would disappear. Unfortunately, the saccade detection algorithm used in this study is still a work in progress and is imperfect. Specifically, it appears that the low frequency power that trails the saccade (appearing in the next window) did not get flagged. We believe this effect was due to the band pass filter. Thus, we were unable to remove all of the low frequency power due to saccades. The nature of this manipulation (in that saccades occurred at a very high frequency in the hard (2 HVTs) conditions) makes interpretation of these EEG results difficult.

4.4 Cardiac Measures

None of the experimental manipulations significantly impacted HR. However, HRV was lower in city conditions (0.0530) than country conditions (0.0542), $F(1,5) = 7.44$, $p < .05$, and

lower in conditions with two HVTs (0.0517) than conditions with one HVT (0.0555), $F(1,5) = 19.46$, $p < .01$. The decrease in HRV during high workload conditions was expected based on previous research (e.g., Roscoe, 1992; Wilson, 1992).

4.5 Eye Measures

Blink duration. Blink duration was shorter in two HVT conditions (0.1005s) than one HVT conditions (0.1099s), $F(1,5) = 13.81$, $p < .05$, and shorter in city conditions (0.1041s) than country conditions (0.1064s), $F(1,5) = 16.77$, $p < .01$.

Blink rate. Participants blinked less in two HVT conditions (16.28) than one HVT conditions (21.65), but this difference was not statistically significant $F(1,5) = 3.98$, $p = .10$. Blink rate was slower in city conditions (18.34) than country conditions (19.59), $F(1,5) = 8.23$, $p < .05$.

Pupil diameter. Pupil diameter was larger during two HVT conditions (4.086mm) than during one HVT conditions (3.837mm), $F(1,5) = 44.33$, $p < .01$. Similarly, pupil diameter was also larger during city routes (4.008mm) than during country routes (3.918mm), although this difference was not significant, $F(1,5) = 3.38$, $p = .125$.

Unexpectedly, pupil diameter was larger during clear conditions (4.057mm) than hazy conditions (3.868mm), $F(1,5) = 14.85$, $p < .05$. It is possible that the pupil light reflex was responsible for this difference. Specifically, we speculated that the haze conditions were actually brighter than the clear conditions because the light reflecting off the haze was brighter than the background (grass, pavement, buildings, etc.). To test this idea we used a Minolta Chroma-Meter CS-100 to assess the luminance of the two conditions. We measured four locations in the simulator four times each and calculated the average. Indeed the average luminance under hazy settings was 59.16 foot-lamberts (fL), whereas the average luminance under clear conditions was 32.61 fL, a difference of 26.55 fL. Thus, the intensity of light was greater in hazy conditions, thereby explaining the constriction in pupil diameter.

Generally, the eye tracking measures followed patterns consistent with previous research. That is, blink duration and blink rate both decreased, and pupil diameter increased in high workload conditions (Wang & Zhou, 2013).

5.0 DISCUSSION

There were two goals of the current research. The first goal was to identify realistic drivers of workload for RPA operators. The second goal was to evaluate the utility of several physiological measures for workload assessment. In future systems we hope to be able to use physiological measures to detect the onset of mental overload, and have tools available to mitigate workload before performance deteriorates. The current research has made an important contribution to this challenge.

In regards to the first goal, the military is continually trying to accomplish more RPA missions using less manpower (Dixon et al., 2004). Advanced RPA systems, such as the one used in this

study, are under development to help meet this goal. We need to examine these systems to understand the potential effects of increased workload on human performance.

According to self-reported workload, two of the experimental manipulations (route type and number of HVTs) were significant drivers of workload. The notion that it is more difficult to track HVTs in congested city areas than open areas was suggested by one of our RPA SMEs. The route manipulation that we implemented confirmed this suggestion, and we were able to identify physiological correlates of increased workload.

In addition, by manipulating the number of RPA the participant controlled, we were able to compare the workload of controlling one vs. two RPA. Not surprisingly, the self-reported workload was significantly higher when controlling two RPA. The control of multiple semi-autonomous air vehicles is not a current capability, and so the present research is valuable in that it provides a preview of what may be expected if/when the capability is implemented. The HVT manipulation also had a strong impact on performance and physiological measures.

The weather manipulation was not a significant driver of workload. This may be due to the nature of the task. Namely, the task was not a visual search task because participants knew where the HVTs were at the beginning of the trial. Also, the HVTs were the only motorcycles in the simulation (other traffic consisted of cars, trucks, vans etc.), so it could be that the haze did not sufficiently obscure the visual cues necessary to track them. We suspect that workload would have been impacted if we had used a denser level of haze, but we were reluctant to do this because we felt that it may have led to total loss of the HVT(s) in too many trials. That is, we could have increased the haze to the point in which the target was no longer visible, or barely visible, but it would have made the task impossible.

The second goal, to evaluate the utility of several physiological measures for workload assessment, was also achieved. That is, the results of this study are promising in that several physiological measures corroborate the story being told by the subjective workload and performance measures. For instance, haze did not impact perceived workload or performance, and it also did not impact HRV, blink duration, or blink rate. Importantly, these physiological measures did detect significant differences when manipulations were effective. Route type and the number of HVTs both significantly impacted perceived workload and performance, and this difference was mirrored by HRV, blink duration, and blink rate. Furthermore, these findings are consistent with previous research (e.g., Wang & Zhou 2013), which provides more confidence that our findings did not occur by chance. This is important because it indicates that these measures are robust and would most likely work in actual RPA field operations.

Despite these promising results, not all physiological measures were sensitive to the workload manipulations. For instance, the HVT and route manipulations did not impact HR (although they did impact HRV). This is consistent with Roscoe (1992), who suggested that HRV may be the superior cardiac measure for assessing cognitive workload.

Additionally, the EEG data is difficult to interpret. Even after accounting for eye artifacts, results were not always in the direction expected. For example, alpha power increased at several sites in two HVT (high workload) conditions. This is in contrast to the classic concept that alpha activity

is an idling rhythm of humans at rest, which becomes desynchronized during cognitive processes (Pfurtscheller & Lopes da Silva, 1999). According to this view, synchronized activity in the alpha band can be interpreted as a neurophysiological correlate of decreased cortical excitability or inhibition of neuronal populations. Thus, we searched for an explanation of our findings, which are not consistent with this perspective. As discussed in the cortical results section, one possible explanation is that EOG artifacts were present in the EEG data due to the additional saccades associated with two target tracking conditions. Importantly, even when examining only EEG windows not co-occurring with saccades, EEG power still appears inflated in the EEG windows trailing saccades because of ringing in the band pass filter. Additionally, the saccade detection algorithm is proficient at finding big saccades, but can miss small saccades. Therefore, all of the saccades associated with the two target tracking condition were not removed.

It is worth noting that other researchers have observed increases in alpha power with high workload as well. Dahlstrom, Nahlinder, Wilson, and Svensson (2011), for instance, observed an increase in alpha power during aerobic flight maneuvers. They suggested that their findings were inconsistent with several studies in which alpha power decreased under high workload (i.e., Dussault et al., 2004; Prinzel et al. 2003; Wilson, 2002), and subsequently advised that their findings could have been a result of muscle artifacts. Thus, interpretation of EEG results can be challenging. Furthermore, meta-analytic results indicate that in general, researchers need to do a better job of conducting and describing their methods of artifact removal (Fatourehchi et al., 2007).

5.1 Limitations

There are limitations in the current study that need to be addressed. First, the sample size is small. At times, measures (e.g., blink rate, pupil diameter) failed to yield significant differences, despite trending in the expected direction. A larger sample size would help to either confirm or deny the utility of these measures.

Another concern in this study is that participants were required to click and hold a button on the right monitor when answering communications. This was highly disruptive to the primary task because it was not possible to actively track HVTs while responding to communications. Based on our debriefing questions, the stress associated with this disruption far exceeded the cognitive component of the questions. In subsequent research, a push-to-talk key on the keyboard will be implemented to eliminate this conflict.

The EEG-based saccade detection algorithm needs improvement. When a saccade occurs, the length of time the band pass filter rings needs to be determined so that the associated EEG data can be flagged as containing an artifact. Secondly, the EEG-based saccades need to be corroborated using EOG-based saccades that report the magnitude and direction of the saccade.

5.2 Implications and Future Research

One important long-term goal of this line of research is to make real-time assessments of operator workload for the purpose of augmenting performance. In the future, researchers should

explore physiologically-based adaptive automation, which is a method of providing assistance to operators by introducing automation only when it is beneficial (Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1996). For example, Wilson and Russell (2007) used physiological features to train an artificial neural network to classify workload, which in turn was used to determine when the operator needed assistance. By monitoring workload and modifying the task demands to match the cognitive capacity of the operator in real-time, the researchers were able to bring about an improvement in performance of approximately 50%. Researchers should continue to evaluate the utility of physiological measures in monitoring workload in a variety of ecologically valid task environments, and explore different means of exploiting this data in real-time to augment operator performance.

6.0 CONCLUSIONS

The current study implemented a novel ecologically valid RPA tracking task, which was used to investigate mental workload, performance, and physiological responses. Two of the three experimental manipulations had a significant effect on the dependent variables. One of the manipulations caused a substantial increase in workload associated with the simultaneous control of multiple RPA. Based on the results of this study, designers of future control stations that allow the control of multiple unmanned vehicles must be cognizant of the substantial increase in workload that can occur as situations dynamically and perhaps unexpectedly become more complex. (e.g., two HVTs emerge from a building instead of one).

Fortunately, results also revealed that multiple physiological measures, which can be obtained continuously and unobtrusively in real time, can provide cues that an operator is facing increased workload. Such cues could be utilized by commanders and/or automated systems to provide augmentation before performance decrements occur. Moreover, it is especially promising that significant physiological differences were observed, even though this was a controlled laboratory task. In actual combat theater where differences in performance can have life or death consequences, these differences will likely be even more distinguishable. For instance, it seems probable that HR would increase more in response to challenges in the field than in the laboratory (Wilson, 1992). In conclusion, physiological measures have potential for real-time workload assessments in RPA task environments, and future research should build on this knowledge for performance augmentation.

7.0 REFERENCES

- Bailey, L. L., & Thompson, R. C. (2001). The TLX: One or more constructs. *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1-4). Columbus, OH: Ohio State University.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292.
- Berka, C., & Levendowski, D. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space and Environmental Medicine*, 78(5), B231-B244.
- Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361-377.
- Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology*, 16, 163-181.
- Cain, B. (2007). *A review of the mental workload literature* (Technical Report M3M 3B9). Toronto, Ontario: Defense Research and Development Canada Toronto.
- Christ, R. E., Hill, S. G., Ayers, J. C., Iavecchia, H. M., Zaklad, A. L., & Bittner, A. (1993). *Application and validation of workload assessment techniques* (Technical Report 974). Alexandria, VA: U.S. Army Research Institute for the Behavioral Sciences.
- Christensen, J. C., & Estepp, J. R. (2013). Coadaptive aiding and automation enhance operator performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55(5), 965-975.
- Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59(1), 57-63.
- Comstock, J. R. & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research* (NASA Tech. Mem. 104174). Hampton, VA: NASA Langley Research Center.
- Cooke, A., Kavussanu, M., McIntyre, D., & Ring, C. (2011). Effects of competition on endurance performance and the underlying psychological and physiological mechanisms. *Biological Psychology*, 86, 370-378.

- Dahlstrom, N., Nahlinder, S., Wilson, G. F., & Svensonn, E. (2011). Recording of psychophysiological data during aerobatic training. *International Journal of Aviation Psychology*, 21(2), 105-122.
- Deci, E. L., & Ryan, R. M. (1991). Intrinsic motivation and self-determination in human behavior. In R. Steers & L. Poerter (Eds.) *Motivation and work behavior* (5th Ed., pp.44-58). New York: McGraw-Hill.
- Dixon, S. R., Wickens, C. D., & Chang, D. (2004). Unmanned aerial vehicle flight control: False alarms versus misses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, New Orleans, LA*, 48(1), 152-156.
- Dussault, C., Jouanin, J. C., & Guézennec, C. Y. (2004). EEG and ECG changes during selected flight sequences. *Aviation Space and Environmental Medicine*, 75(10), 889-897.
- Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3), 480-494.
- Festinger, L. (1954). A theory of social comparison process. *Human Relations*, 7, 117-140.
- Finomore, V., Popik, D., Dallman, R., Stewart, J., Satterfield, K., Castle, C. (2011). Demonstration of a Network-Centric Communication Management Suite: Multi-Modal Communication. In *Proceedings of the 55th Human Factors and Ergonomics Society Annual Meeting*, HFES, Las Vegas.
- Gevins, A. S., DuRousseau, D., Zhang, J., & Libove, J. (1993). *Flight helmet EEG system* (Final Tech. Report AL/CF-SR-1993-0007). San Francisco, CA: Sam Technology.
- Gevins, A., S., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7(4), 374-385.
- Graham, S., Jiang, J., Manning, V., Nejad, A. B., Zhisheng, K., Salleh, S. R., et al. (2010). IQ related fMRI differences during cognitive set shifting. *Cerebral Cortex*, 20(3), 641-649.
- Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., & Rao, R. P. N. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *CHI Conference on Human Factors in Computing Systems*. Florence, Italy.
- rossman, P. (1992). Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: Selection of window frames, keeping the panes clean and viewing the neural topography. *Biological Psychology*, 34(2-3), 131-161.
- Hamm, J. P., Dyckman, K. A., Ethridge, L. E., Mcdowell, J. E., & Clementz, B. A. (2010). Preparatory activations across a distributed cortical network determine production of express saccades in humans. *Journal of Neuroscience*, 30(21), 7350-7357.

- Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, 69, 360-367.
- Harrison, L. K., Denning, S., Easton, H. L., Hall, J. C., Burns, V. E., Ring, C., Carroll, D. (2001). The effects of competition and competitiveness on cardiovascular activity. *Psychophysiology*, 38, 601-606.
- Hart, S. G., & Staveland, L. E. (1988). Development of the NASA- TLX (Task Load Index): Results of experimental and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.138-183). Amsterdam: North-Holland Press.
- Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35, 579-601.
- Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. In *Conference on Human Factors in Computing Systems* (pp. 1477-1480). New York: ACM Press.
- Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9(3), 152-169.
- Jasper, H. (1958). Report of the committee on methods of clinical examination. *Electroencephalography and Clinical Neurophysiology*, 10, 370-375.
- Jorna, P. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, 34(2-3), 237-257.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352-358.
- Kramer, A. F. (1991). Physiological measures of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple-task performance* (pp. 279-238). London: Taylor & Francis.
- Liu, D., Wasson, R., & Vincenzi, D. A. (2009). Effects of system automation management strategies and multi-mission operator-to-vehicle ratio on operator performance in UAV systems. *Journal of Intelligent and Robotic Systems*, 54(5), 795-810.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2), 1-13.

- Marshall, S. P., Pleydell-Pearce, C. W., & Dickson, B. T., (2003). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. *Proceedings of the 36th Hawaii International Conference on System Sciences, Hawaii*, 130-135.
- Martens, R. (1975). *Social psychology and physical activity*. New York: Harper and Row.
- McFarland, D. J., McCane, L. M., David, S.V., Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*, 103, 386-394.
- Moroney, W. F., Biers, D. W., & Eggemeier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload measurement techniques. *The International Journal of Aviation Psychology*, 5(1), 87-106.
- Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34(2-3), 205-236.
- Müller-Plath, G. (2008). Localizing subprocesses of visual search by correlating local brain activation in fMRI with response time model parameters. *Journal of Neuroscience Methods*, 171(2), 316-330.
- Niedermeyer E., & da Silva F.L. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincot Williams & Wilkins.
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17-31.
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance: Vol. 2, Cognitive Processes and Performance* (pp. 42.1-42.49). New York: Wiley.
- Parasuraman, R., Mouloua, M. & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(4), 665-679.
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110, 1842-1857.
- Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. In D. Harris, V. Duffy, M. Smith, & C. Stephanidis (eds.), *Human-Centred Computing: Cognitive, Social, and Ergonomic Aspects*. Vol. 3 of the *Proceedings of the 10th International Conference on Human-Computer Interaction, HCI 2003*, Crete, Greece.

- Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology*, 34(2-3), 93-130.
- Pouliot, P., Tremblay, J., Robert, M., Vannasing, P., Lepore, F., Lassonde, M., et al. (2012). Nonlinear hemodynamic responses in human epilepsy: A multimodal analysis with fNIRS-EEG and fMRI-EEG. *Journal of Neuroscience Methods*, 204(2), 326-340.
- Prabhala, S. V., Gallimore, J. J., & Narayanan, S. (2003). Emerging areas: Urban operations and UCAVs: Human effectiveness issues in simulated uninhabited combat aerial vehicles. In *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation* (pp. 1034-1038). New Orleans, LA.
- Prinzel, L. J., Parasuraman, R., Freeman, F. G., Scerbo, M. W., Mikulka, P. J. & Pope, A. T. (2003). *Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation design* (Tech. Rep. No. 2003-212 442). Hampton, VA: NASA Langley Research Center.
- Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, 34(2-3), 259-287.
- Rose, M. R., Arnold, R. D., & Howse, W. R. (2013). Unmanned aircraft systems selection practices: Current research and future directions. *Military Psychology*, 25(5), 413-427.
- Rowe, A. J., Liggett, K. K., Davis, J. E. (2009). Vigilant Spirit Control Station: A Research Testbed for Multi-UAS Supervisory Control Interfaces. In *Proceedings of the 15th International Symposium on Aviation Psychology*, Dayton, OH.
- Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators and Virtual Environments*, 11(4), 335-351.
- Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* [CD-ROM]. Mahwah, NJ: Erlbaum.
- Seta, J. J. (1982). The impact of comparison processes on coactors' task performance. *Journal of Personality and Social Psychology*, 42, 281-291.
- Sherif, M., Harvey, O. J., White, B. J., Wood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robber's cave experiment*. Normal, OK: Institute of Group Studies.
- Sommer, S. (1995). Social competition: Identifying new perspectives and strategies for task motivation. *The International Journal of Conflict Management*, 6 (3), 239-256.

- Tsang, P. & Wilson, G. (1997). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 243-273). New York: John Wiley.
- U.S. Department of Defense. (2011). *Unmanned systems integrated roadmap FY2011-2036* (Reference No. 11-S-3613). Washington, DC: Department of Defense.
- Veldhuijzen van Zanten, J. J. C. S., De Boer, D., Harrison, L. K., Ring, C., Carroll, D., Willemsen, G., et al., 2002. Competitiveness and hemodynamic reactions to competition. *Psychophysiology*, 39, 759-766.
- Veltman, J. A., & Gaillard, A. W. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5), 656-669.
- Wang, Y. & Zhou, J. (2013). Literature review on physiological measures of cognitive workload. Machine learning research group – NICTA.
- Wickens, C. (1984). Processing resources in attention. In R. Parasuraman & D. Davies (Eds.), *Varieties of Attention* (pp. 63-102). Orlando: Academic Press.
- Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, 34(2-3), 163-178.
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.
- Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. *The International Journal of Aviation Psychology*, 17(2), 219-247.
- Wilson, G. F. & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 329-360). London: Taylor & Francis.
- Wilson, G., et al. (2004). *Operator functional state assessment*. Research and Technology Organization, North Atlantic Treaty Organization (NATO).
- Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(6), 1005-1018.
- Wolpaw, J. R., Mcfarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3), 252-259.
- Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, 48(1), 280-290.

- Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, 3(2), 178-194.
- Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2), 1-5.
- Zarjam, P., Epps, J., & Lovell, N. H. (2012). Characterizing mental load in an arithmetic task use entropy-based features. In *The 11th International Conference on Information Sciences, Signal Processing, and their Applications: Main Tracks* (pp.199-204).

APPENDIX A-Screenshots



Figure A-1. Condition 1 (One Target, Country, Sunny).



Figure A-2. Condition 2 (One Target, City, Sunny).



Figure A-3. Condition 3 (Two Targets, Country, Sunny).



Figure A-4. Condition 4 (Two Targets, City, Sunny). The video feed on the left is from the second highest (6.9x) level of zoom, and the video feed on the right is from the third highest (4.4x) level of zoom.



Figure A-5. Condition 5 (One Target, Country, Hazy).



Figure A-6. Condition 6 (One Target, City, Hazy).



Figure A-7. Condition 7 (Two Targets, Country, Hazy).



Figure A-8. Condition 8 (Two Targets, City, Hazy).



Figure A-9. Overall Display (includes VSCS on left and middle monitors and MMC on the right monitor).



Figure A-10. Tactical Situation Display.

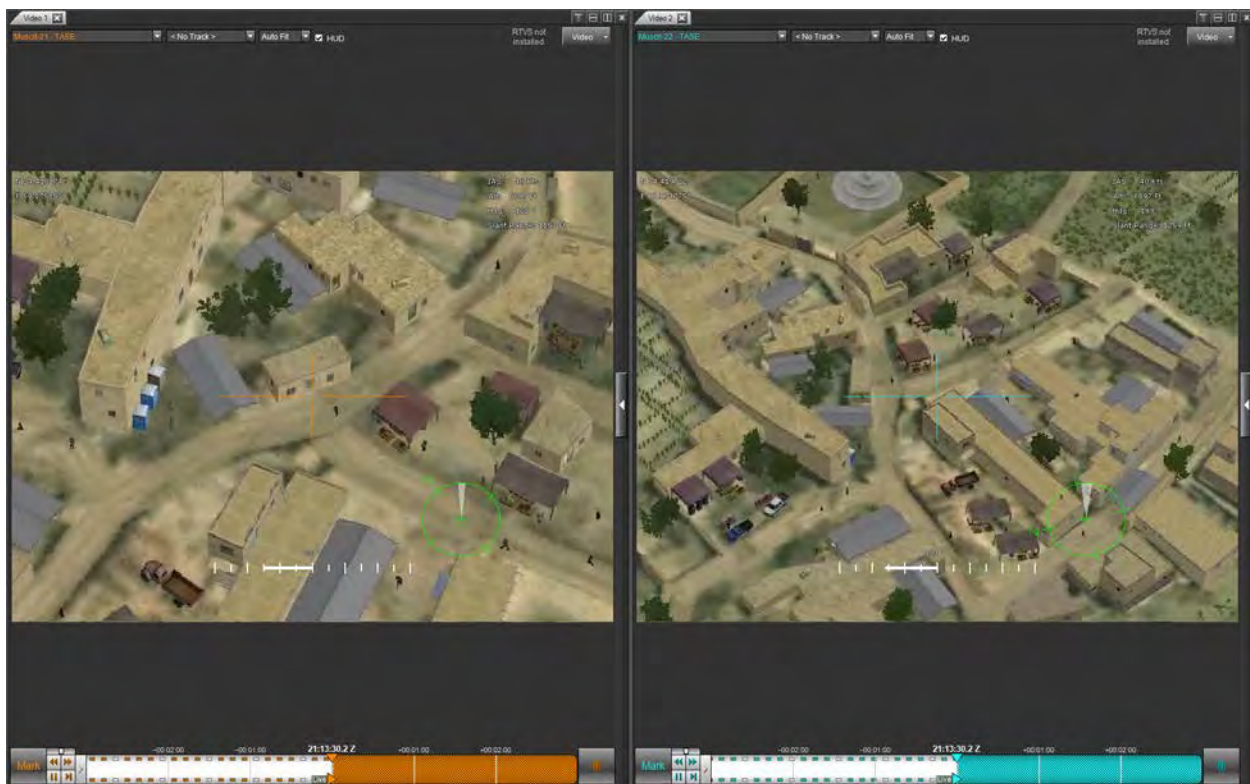


Figure A-11. Sensor feeds. RPA 1 sensor feed is on the left, and RPA 2 sensor feed is on the right.



Figure A-12. Multi-Modal Communication.

APPENDIX B-EEG Reference

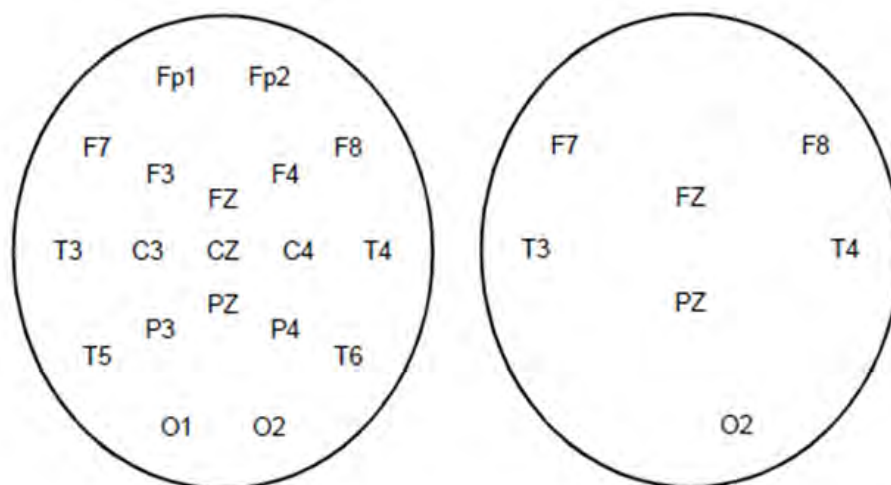


Figure B-1. All EEG site locations from the International 10/20 system (shown on left) vs. site locations used in the current study (shown on right).

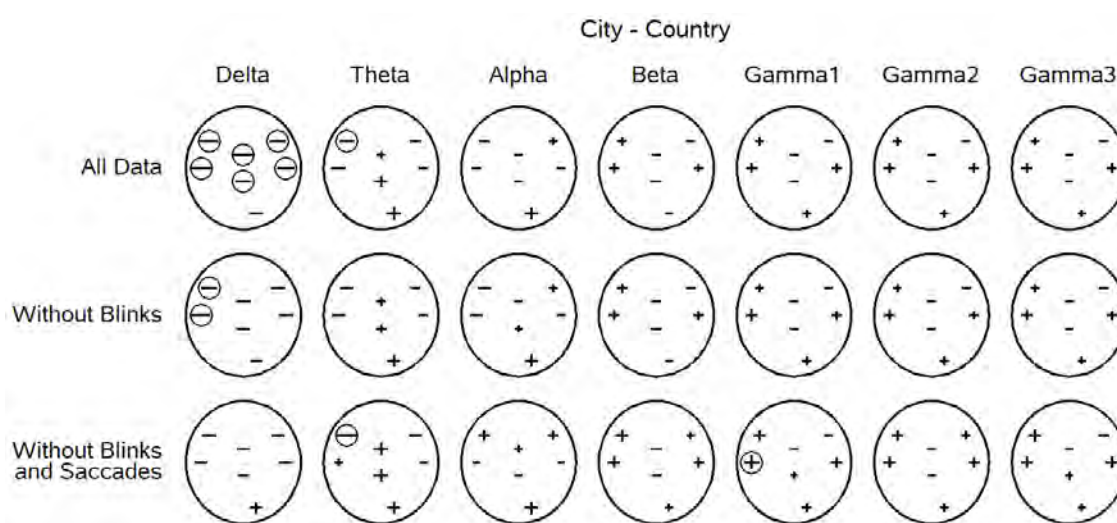


Figure B-2. EEG power for the route manipulation. The sign is direction of the difference in log power (for example, a plus sign means more power for city than country). The size of the sign is relative absolute value of the t statistic (i.e., larger sign means smaller p-value). If sign is circled then $p \leq 0.05$.

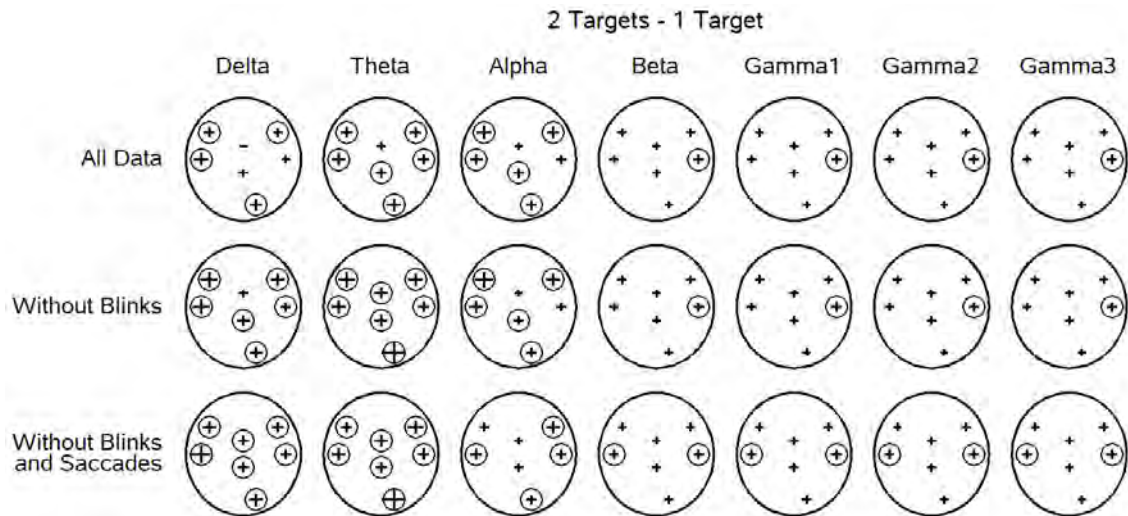


Figure B-3. EEG power for the target manipulation. The sign is direction of the difference in log power (for example, a plus sign means more power for 2 targets than 1 target). The size of the sign is relative absolute value of the t statistic (i.e., larger sign means smaller p-value). If sign is circled then $p \leq 0.05$.

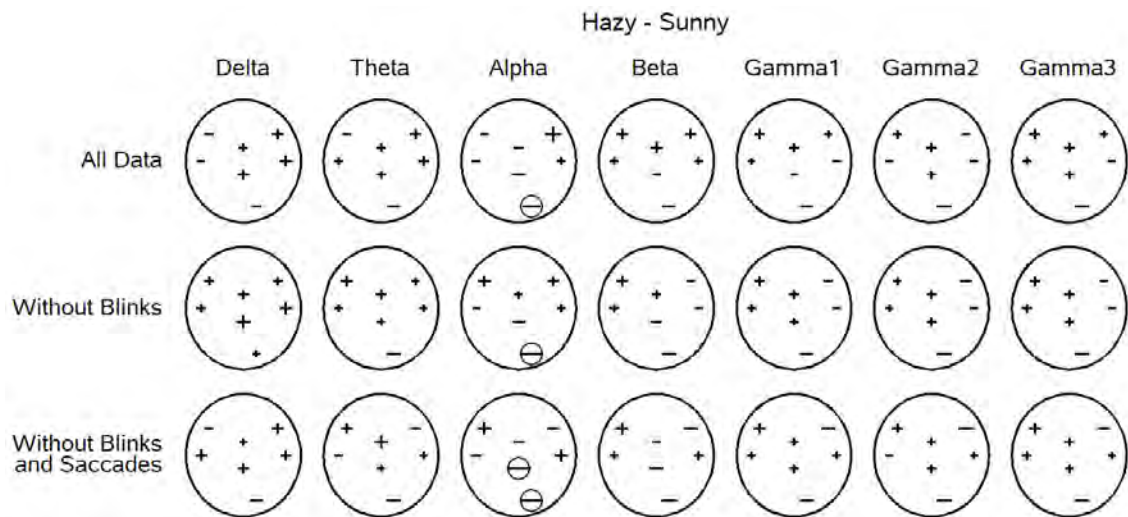


Figure B-4. EEG power for the weather manipulation. The sign is direction of the difference in log power (for example, a plus sign means more power for hazy than sunny). The size of the sign is relative absolute value of the t statistic (i.e., larger sign means smaller p-value). If sign is circled then $p \leq 0.05$.

LIST OF ABBREVIATIONS AND ACRONYMS

ANOVA	analysis of variance
ECG	electrocardiography
EEG	electroencephalography
EOG	electrooculography
fMRI	functional magnetic resonance imaging
HR	heart rate
HRV	heart rate variability
HUD	heads-up display
HUMAN	human universal measurement and assessment network
HVT	high value target
Hz	Hertz
MMC	Multi-Modal Communication
MRT	multiple resource theory
RPA	remotely piloted aircraft
SE	standard error
SME	subject matter expert
TLX	Task Load Index
VEOG	vertical electrooculography
VRSG	Virtual Reality Scene Generator